PLOS PATHOGENS

RESEARCH ARTICLE

Genomic analysis reveals an exogenous viral symbiont with dual functionality in parasitoid wasps and their hosts

Kelsey A. Coffman [ID]*, **Gaelen R. Burke** [ID]

Department of Entomology, University of Georgia, Athens, Georgia, United States of America

* kcoffman@uga.edu

Abstract

Insects are known to host a wide variety of beneficial microbes that are fundamental to many aspects of their biology and have substantially shaped their evolution. Notably, parasitoid wasps have repeatedly evolved beneficial associations with viruses that enable developing wasps to survive as parasites that feed from other insects. Ongoing genomic sequencing efforts have revealed that most of these virus-derived entities are fully integrated into the genomes of parasitoid wasp lineages, representing endogenous viral elements (EVEs) that retain the ability to produce virus or virus-like particles within wasp reproductive tissues. All documented parasitoid EVEs have undergone similar genomic rearrangements compared to their viral ancestors characterized by viral genes scattered across wasp genomes and specific viral gene losses. The recurrent presence of viral endogenization and genomic reorganization in beneficial virus systems identified to date suggest that these features are crucial to forming heritable alliances between parasitoid wasps and viruses. Here, our genomic characterization of a mutualistic poxvirus associated with the wasp *Diachasmimorpha longicaudata*, known as Diachasmimorpha longicaudata entomopoxvirus (DIEPV), has uncovered the first instance of beneficial virus evolution that does not conform to the genomic architecture shared by parasitoid EVEs with which it displays evolutionary convergence. Rather, DIEPV retains the exogenous viral genome of its poxvirus ancestor and the majority of conserved poxvirus core genes. Additional comparative analyses indicate that DIEPV is related to a fly pathogen and contains a novel gene expansion that may be adaptive to its symbiotic role. Finally, differential expression analysis during virus replication in wasps and fly hosts demonstrates a unique mechanism of functional partitioning that allows DIEPV to persist within and provide benefit to its parasitoid wasp host.

Author summary

Viruses have repeatedly formed long-term associations with insects called parasitoid wasps, which grow as parasites within other insect hosts. While these viruses were once pathogenic, they have since been co-opted by parasitoid wasps to benefit the survival of wasp offspring during parasitism. The genomes of most identified beneficial viruses are

fully integrated into the genomes of the parasitoid wasps that produce them. Because
these virus-derived entities have lost the ability to exist apart from their associated wasps,
they are considered endogenous viral elements (EVEs) of the wasps rather than mutualis-
tic symbionts. We sequenced the genome of the beneficial parasitoid virus Diachasmimor-
pha longicaudata entomopoxvirus (DlEPV) and found that its genome is not integrated
into the genome of *Diachasmimorpha longicaudata* wasps and has largely retained the
genomic structure of its pathogenic ancestor. Given the importance of viral genome inte-
gration in the overall stability of parasitoid wasp-EVE systems, we identified a novel strat-
egy used by DlEPV to maintain its relationship with *D. longicaudata* despite its lack of
endogenization. Our findings in this study demonstrate the first instance of a mutualistic
viral symbiont in insects and provide new insight into the means through which beneficial
viruses can arise.

Introduction

Microbial symbionts have been increasingly identified as major drivers of animal evolution
due to the novel capabilities microbes provide to their hosts and the speed at which symbiosis
can cause adaptive change in animal lineages [1,2]. Insects, in particular, have repeatedly
formed symbiotic alliances with microbes that highly vary with respect to taxonomic classifica-
tion, localization on or within the insect, transmission strategies, and phenotypic traits pro-
vided to the insect [3]. Bacterial symbionts have been the primary focus of study for many
insect groups, such as plant sap-feeders, blood-feeders, and social insects [4–6]. However, par-
asitoid wasps, whose young are obligate parasites of other arthropods, are better known for
their numerous associations with viruses [7–9]. Parasitoid wasp lineages have repeatedly
acquired heritable viruses in conjunction with evolutionary arms race dynamics between
wasps and their hosts [10–12]. Many of these associations are extraordinary examples of
endogenous viral elements (EVEs) within wasp genomes, in which components of viral
machinery are retained from their pathogenic ancestors to produce virus or virus-like particles
within wasp ovaries [13,14]. The resulting virus-derived particles accompany wasp eggs when
delivered into host insects and can function to protect parasitoid eggs from attack by the host
immune system and/or actively disrupt host developmental and immunological pathways
[10,15,16].

Rather than existing in a wasp genome as a contiguous region of proviral DNA, parasitoid
EVEs share an unconventional genomic architecture characterized by the dispersal of viral
genes to separate regions of the wasp genome [17–22]. Key virus replication genes have also
been lost in all cases for which the viral ancestor is known, which implies that wasp genes are
instead needed to complete virus particle production [23]. These genomic anomalies have
three major consequences that are thought to be important adaptations in parasitoid-EVE
associations. First, permanent integration of the viral genome into the wasp genome ensures
viral transmission to future wasp generations. Second, viral gene dispersal and gene loss for-
feits EVE autonomy over their own propagation, allowing for strict regulation of virus replica-
tion by the wasp. Third, virus particles produced in wasp tissue do not contain the necessary
genes for further replication outside of the wasp. The functional outcome of these genomic fea-
tures is most clearly understood in the polydnaviruses (PDVs), a group of ancient EVEs
formed from multiple, independent viral acquisition events [17,18]. Due to their unusual
genome organization, PDVs contain a dual functionality that is effectively split between two
insects: PDV replication occurs exclusively in wasp tissue, while PDV virulence is confined to

parasitized host tissue [24,25]. This separation of virus function promotes stability within these associations, because it minimizes wasp-virus conflict and establishes an interdependency for survival, in which wasp offspring depend on PDV virulence within the host, and PDVs depend on wasps for transmission and amplification [26,27]. Furthermore, the recurrent observation of this distinctive genomic architecture in more recently acquired EVEs supports the notion that viral genome integration and reorganization is fundamental to the persistence of wasp-virus mergers [21,22].

However, additional examples of heritable viruses have been identified in parasitoid lineages that are of unique viral origin and may deviate from this pattern. An ascovirus carried by the wasp *Diadromus pulchellus*, named Diadromus pulchellus toursvirus (DpTV), contains a circular, episomal DNA genome present in the nuclei of wasp cells, and when DpTV virions infect the caterpillar hosts of *D. pulchellus*, the virus inhibits the host melanization response during an immune challenge [28–30]. Additionally, a heritable iflavirus discovered within *Dinocampus coccinellae* parasitoid wasps, known as Dinocampus coccinellae paralysis virus (DcPV), contains an exogenous RNA viral genome that replicates within the neural tissue of coccinellid beetle hosts of the wasps during parasitism. This viral activity is thought to cause a behavioral manipulation within the host, in which parasitized beetles will guard the parasitoid pupa against predation [31]. While these examples provide suggestive evidence that the virus in each case is beneficial for the parasitoid that transmits it, neither example has been experimentally shown to provide a direct fitness benefit for its associated wasp. An effective method to determine whether heritable viruses are truly mutualistic for wasps is to remove the virus population from wasps and compare the success of "cured" wasps to those with a normal viral load. Lower survivorship of wasp progeny when not accompanied by virus is strong evidence that the virus provides a net benefit to wasp fitness.

This has been recently demonstrated for *Diachasmimorpha longicaudata* parasitoid wasps and the heritable poxvirus female wasps maintain within their venom gland, known as Diachasmimorpha longicaudata entomopoxvirus (DlEPV) [32,33]. We showed that DlEPV is vertically transmitted to each wasp generation within oviposited wasp eggs and provides a considerable boost to wasp survival during parasitism within *Anastrepha suspensa* fruit fly hosts, because wasps reared without DlEPV survive at a drastically reduced rate compared to wasps with a typical viral load [34]. DlEPV is currently the only mutualistic poxvirus to be identified in parasitoid wasps, and unlike EVEs, DlEPV can replicate within both wasps and fruit fly hosts of the wasps [34]. Therefore, DlEPV appears to have replicative autonomy within both insects, suggesting that this virus retains more features from its pathogenic ancestor than other parasitoid viral elements. Despite these ancestral characteristics, we have also demonstrated that DlEPV replication is highly virulent within host fly tissue, while replication within wasp tissue has no observable pathogenic effects [34]. These results imply that DlEPV utilizes a similar strategy of functional partitioning to that observed in parasitoid EVEs. DlEPV is therefore unique in that it maintains features of both an autonomous viral pathogen and a beneficial viral symbiont. In this study, we sequenced the complete DlEPV genome to ascertain whether DlEPV shares the genomic architecture of parasitoid EVEs and to determine how DlEPV has evolved in comparison to other poxviruses, including the identification of its closest known relative. Using the results from our comparative analyses, we then performed a functional genomic investigation to elucidate the novel means through which DlEPV may achieve its beneficial relationship with *D. longicaudata*.

Results

Sequencing and assembly of the DlEPV genome

The DlEPV genome is non-endogenous. Poxviruses are large DNA viruses that infect vertebrates (chordopoxviruses, or CPVs), as well as insects (entomopoxviruses, or EPVs) [35].

The study of poxviruses has historically focused on CPVs and the prototype CPV, known as vaccinia virus (VACV), due to the societal impact of smallpox [36]. EPVs have been comparatively neglected but function similarly to CPVs in many ways, while exhibiting differences that can often be attributed to the biology of their insect hosts [37]. Both CPV and EPV genomes exist as linear, double-stranded DNA (dsDNA) molecules that contain a hairpin loop at each terminus. The two extreme ends of the genome consist of sequence repeats that are inversions of one another, known as inverted terminal repeats (ITRs), while the genome interior contains the majority of viral genes [38]. The DlEPV genome sequence, obtained from high-throughput sequencing of *D. longicaudata* venom gland DNA, was assembled into a single 253 kilobase (kb) contiguous sequence, including two 17 kb ITR regions and 193 open reading frames (ORFs) (Fig 1 and S1 Table). The contiguity and lack of flanking wasp genes in our assembly implies that the DlEPV genome is not endogenous within the wasp genome. Normalized quantitative PCR (qPCR) data of viral abundance in wasp tissue also support this finding by showing that the number of DlEPV genome copies is less than the number of wasp genome copies for several tissues, developmental stages, and all male wasps (S1 Fig).

The DlEPV genome has abnormally low coding density and high GC content.   The majority of publicly available EPV genomes are from lepidopteran (moth and butterfly) poxviruses: Amsacta moorei entomopoxvirus (AMEV), Adoxophyes honmai entomopoxvirus (AHEV), Choristoneura biennis entomopoxvirus (CBEV), Choristoneura rosaceana entomopoxvirus (CREV), and Mythimna separata entomopoxvirus (MySEV) [39,40]. Orthopteran (grasshopper) and coleopteran (beetle) poxvirus genomes contain single representatives: Melanoplus sanguinipes entomopoxvirus (MSEV) and Anomala cuprea entomopoxvirus (ACEV), respectively [41,42]. Recently, two additional EPV sequences have been reported. A partial poxvirus genome sequence identified within the argentine ant, named Linepithema humile entomopoxvirus 1 (LHEV), represents the first hymenopteran (ant, wasp, and bee) poxvirus to be sequenced [43]. In addition, a complete poxvirus genome obtained from *Drosophila melanogaster*, known as Yalta virus, represents the first sequenced dipteran (fly) poxvirus [44].

DlEPV, in comparison to these other EPVs, has a similar overall genome length, ITR length, and ORF number (S2 Table). However, the DlEPV genome is peculiar with respect to its coding capacity and GC content. DlEPV is extremely gene-sparse relative to its genome size and contains a heavily reduced coding density of 65.1% compared to the 89.9% ± 3.0% SD coding density of other EPV genomes (S2 Table). This reduced gene density is an exception to poxvirus genomes, in general, which are highly compact with a dense array of non-overlapping genes [38]. The nucleotide composition of the DlEPV genome also varies compared to other EPV genomes, which consistently exhibit the most severe AT-bias found in the poxvirus family [45]. The GC content of the DlEPV genome at 30.1% is substantially higher than the average 20.5% ± 2.4% SD of its EPV relatives (S2 Table). Since viral GC content can be correlated to the GC of the host genome [46], we also estimated *D. longicaudata* and *A. suspensa* genome nucleotide composition using transcriptomes produced for a subsequent differential expression analysis (see Functional Genomic Analysis of DlEPV). Assembled *D. longicaudata* transcripts had 40.7% GC overall, and *A. suspensa* fly hosts contained transcripts with a GC content of 39.7%.

DlEPV genome annotation

DlEPV contains most poxvirus core genes.   We next annotated the DlEPV genome to assess its completeness compared to other poxviruses. The central region of the linear poxvirus genome generally contains genes that are required for virus replication, including the 49 core genes conserved among all sequenced poxviruses [45,47,48]. We were able to identify the majority of poxvirus core genes in the DlEPV genome, with the exception of the following
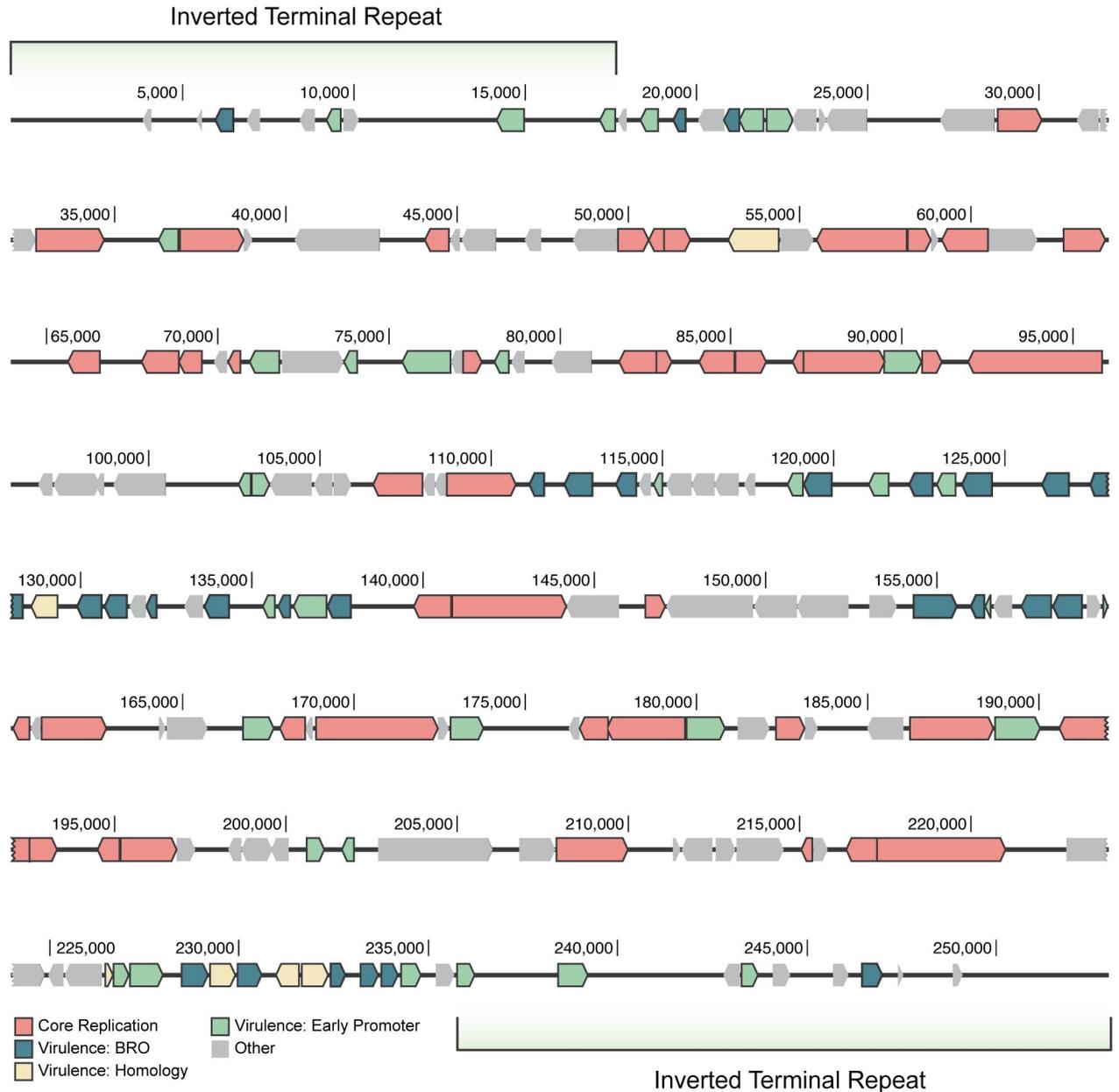
Fig 1. Linear map of the DlEPV genome. Each arrow indicates the genomic position of a DlEPV ORF, and the direction of the arrow corresponds to its strand orientation. Arrows are colored based on the putative functional category of each ORF as defined in the legend at the bottom of the map. Core Replication refers to the 45 poxvirus core genes identified in the DlEPV genome. Virulence: BRO refers to the 27 DlEPV BRO genes, Virulence: Homology indicates the 6 ORFs with similarity to known virulence genes, and Virulence: Early Promoter are the additional 34 putative virulence genes based on the presence of the conserved EPV early promoter sequence and no other assigned function.

https://doi.org/10.1371/journal.ppat.1009069.g001

four genes: the heparin binding surface protein (VACV core gene H3L), a virion core protein (E6R), the NlpC/P60 superfamily protein (G6R), and a RNA polymerase subunit (A29L) [49] (S1 Table). VACV core genes H3L and E6R are both required for the correct assemblage of mature virions, a process known as morphogenesis [50–54]. G6R is unique among the poxvirus core gene set, as its protein product is not required for VACV replication *in vitro* but instead is involved in virulence [55]. A29L encodes the 35 kDa RNA polymerase subunit

(RPO35), one of five conserved subunits of the poxvirus RNA polymerase holoenzyme responsible for viral gene transcription [56]. We utilized our previously reported transcriptome of the *D. longicaudata* venom gland [34] to determine whether these four genes had been transferred from the DlEPV genome to the *D. longicaudata* genome. Endogenized PDV replication genes were first identified in PDV-producing wasps using transcriptome sequencing of wasp ovary tissue collected during PDV replication [17,18]. We therefore hypothesized that DlEPV transcripts with sequence similarity to the undetected genes would be present during virus replication in the venom gland if these genes were endogenous. However, our transcriptome searches yielded no hits to the aforementioned genes.

DlEPV is most closely related to a *Drosophila* poxvirus. Due to the exogenous state of the DlEPV genome and its relatively complete set of core genes, DlEPV appears to be more biologically similar to its viral progenitor than has been observed of parasitoid EVEs. This level of genomic preservation led us to investigate the origin of DlEPV among other poxviruses through phylogenetic reconstruction and identification of its closest relative. Because DlEPV replicates in both a parasitoid wasp and the wasp's fruit fly hosts, it is likely that DlEPV originated as either a parasitoid pathogen or as a fly pathogen. The two most recently published EPV genomes, LHEV and Yalta virus, could therefore give more context on the origin of DlEPV.

We generated a maximum likelihood (ML) phylogeny using 16 concatenated poxvirus core genes from all sequenced EPVs and the following CPVs to test the two hypotheses: VACV, orf virus (ORFV), molluscum contagiosum virus (MOCV), fowlpox virus (FWPV), crocodilepox virus (CRV), and salmon gill poxvirus (SGPV) (Fig 2 and S3 Table). The placement of the 7 originally sequenced EPVs (MSEV, AMEV, AHEV, MySEV, CREV, CBEV, and ACEV) on the tree shows concordance with the higher phylogenetic relationships of their insect hosts, which is consistent with previous EPV phylogenetic analyses [37,40]. In contrast, LHEV and Yalta virus show a clear divergence from other EPVs [44]. The inclusion of DlEPV in our phylogeny revealed that it shares a more recent common ancestor with Yalta virus than LHEV, suggesting that DlEPV is more likely derived from a fly pathogen rather than a parasitoid pathogen (Fig 2). The shared common ancestry between DlEPV and Yalta virus was maintained in an expanded 44 core gene phylogeny that excluded the partial LHEV genome, and the overall tree topology was robust to Bayesian inference analysis (S2 Fig). While the position of the DlEPV/Yalta virus clade appears to bridge the gap between EPVs and CPVs in the unrooted phylogeny, the inclusion of members of the sister group to poxviruses, known as Asfarviridae [57], as an outgroup in a 10 poxvirus core gene phylogeny confirmed that DlEPV and Yalta virus are, in fact, more closely related to EPVs than to CPVs (S2 Fig).

We further investigated similarities between the DlEPV and Yalta virus genomes by searching for additional orthologous genes shared among them. We determined that 44 of the 49 poxvirus core genes are shared between the DlEPV and Yalta virus genomes (S3 Table). Interestingly, 2 of the 3 "missing" genes in the Yalta virus genome were also not detected in the DlEPV genome [44]. This suggests that the absence, or more likely, extreme sequence divergence of these genes is lineage-specific to fly poxviruses, rather than due to genome incompleteness. In addition to the 44 core genes shared between the two genomes, we found 24 single-copy orthologs and 3 orthologs that had undergone duplication in either genome (S1 Table). Most of these orthologous groups are of unknown function, while some have putative functions that are not found in other EPVs and therefore, may be unique to fly poxviruses. These include a ribonucleotide reductase large subunit (DLEV028/Yalta121), an alpha/beta fold hydrolase (DLEV038/Yalta165), and a type II topoisomerase (DLEV158/Yalta014).

While many CPV genomes have a highly conserved gene order [58], this colinear pattern does not hold true for EPVs, which display little synteny with CPVs or EPVs from separate
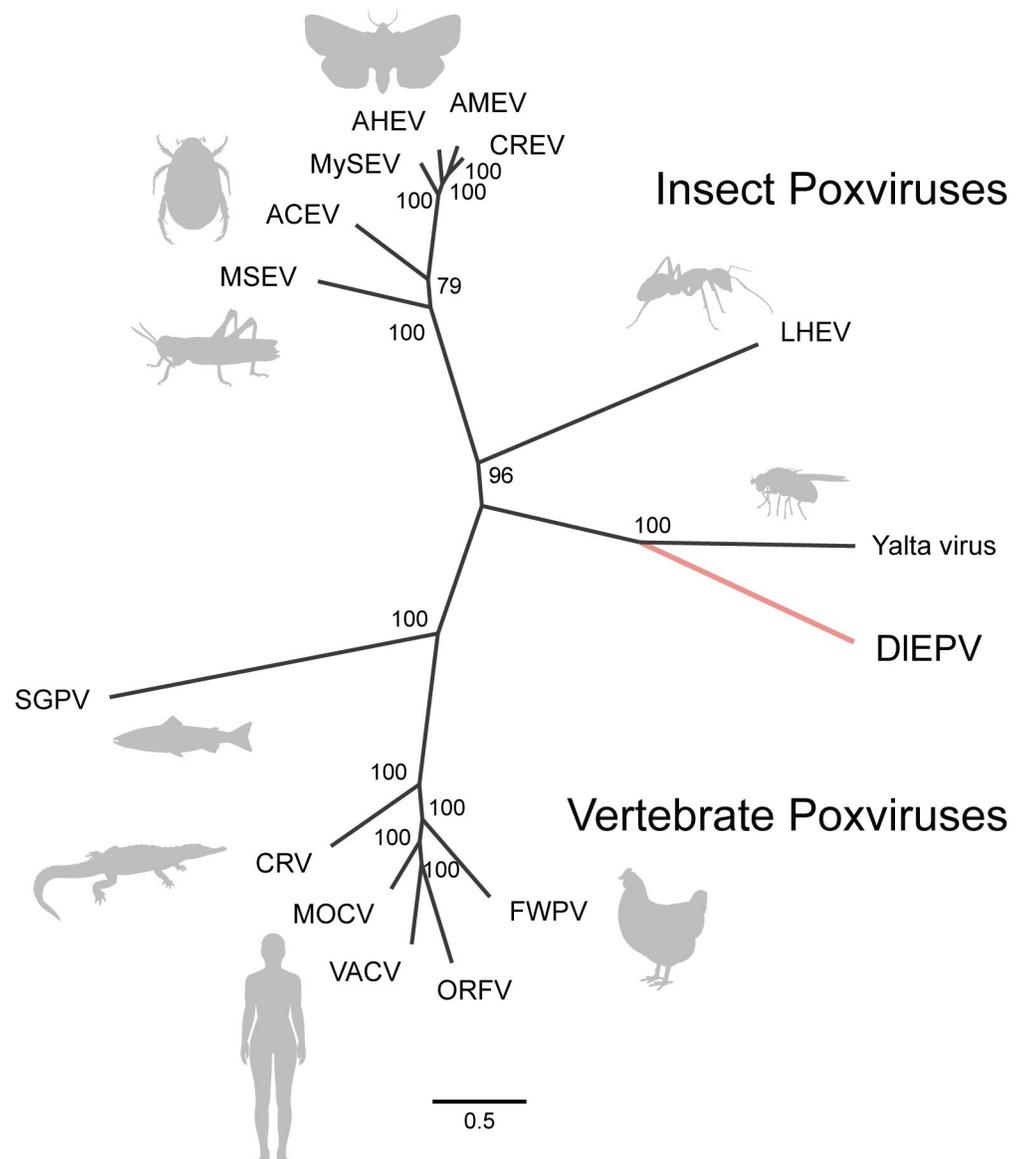
Fig 2. Poxvirus core gene phylogeny demonstrates that the fly-infecting Yalta virus is the closest known relative of DlEPV. Phylogenetic tree constructed from a maximum likelihood analysis using the concatenated amino acid multiple sequence alignment from 16 conserved poxvirus core genes. Node support (%) was inferred with 1,000 bootstrap iterations. Insect and vertebrate poxvirus orthologs used to build the phylogeny are indicated in S3 Table. Genome abbreviations are as defined in the Results section, and accessions are included in the Materials and Methods section.

https://doi.org/10.1371/journal.ppat.1009069.g002

host genera [39,40,42]. We investigated genome synteny between DlEPV and Yalta virus by generating two-dimensional dot plots comparing the genomic positions of their shared 44 core genes (Fig 3). The DlEPV-Yalta virus dot plot revealed a moderate amount of synteny between the two viral genomes. In particular, a large syntenic region of approximately 50 kb was identified, as indicated by the negative linear arrangement of orthologs in the lower-right quadrant of the plot (Fig 3A). This partial synteny further supports a closer relationship between DlEPV and Yalta virus, because both genomes have relatively low synteny when compared to the next closest relative MSEV (Fig 3B and 3C).
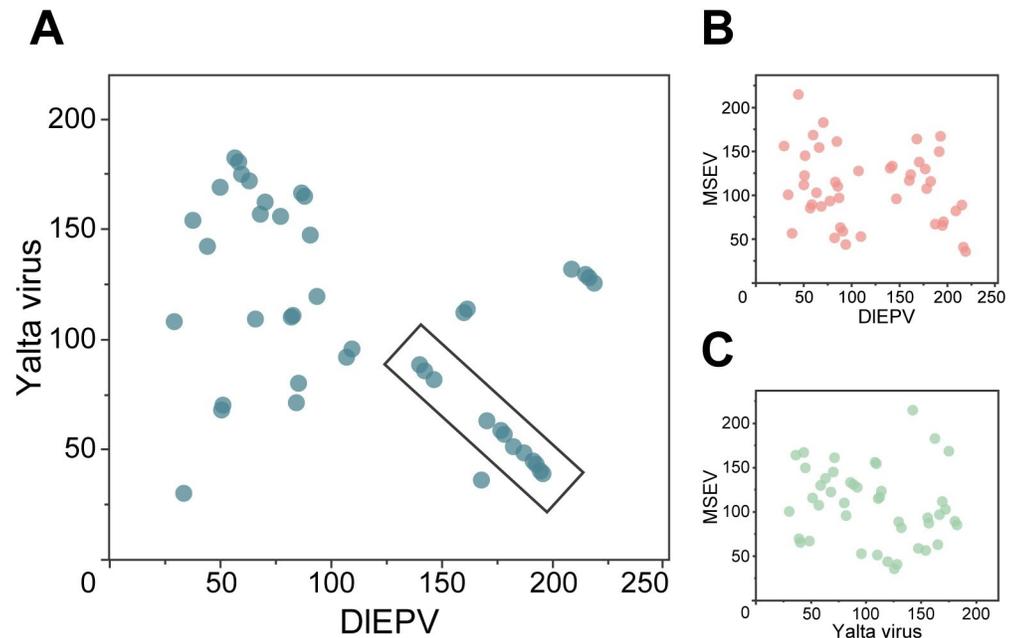
Fig 3. Core gene synteny further supports the close relationship between DlEPV and Yalta virus. Dot plots show the relative genomic location of the 44 poxvirus core genes shared between the DlEPV and Yalta virus genomes when compared to (A) one another, or (B-C) when either is compared to MSEV. Each dot represents a homologous core gene, and axes indicate the genomic position in kilobases. Box in panel A highlights a highly syntenic region between DlEPV and Yalta virus core genes.

The DlEPV genome contains a novel BRO gene expansion. In contrast to the more conserved center of the typical poxvirus genome, the exterior regions contain a variable assortment of virulence genes, which are involved in host interactions [45,59]. Multigene families are common components of large DNA viral genomes that likely represent lineage-specific adaptations resulting from coevolution between the virus and its host [60]. One such family, known as baculovirus repeated ORF (BRO) proteins, are found in many insect DNA viruses, including baculoviruses, iridoviruses, ascoviruses, and EPVs [61]. The function of BRO proteins remains unclear, although they localize to the nuclei of insect cells infected with the baculovirus Bombyx mori nucleopolyhedrovirus [62], and the characteristic N-terminal BRO domain can bind DNA [63]. Therefore, it has been proposed that BRO proteins may play a role in virulence through the regulation of host DNA transcription and/or replication [63]. BRO genes are consistently found in the terminal regions of EPV genomes amongst other virulence genes [39–42]. Strikingly, the DlEPV genome contains 27 BRO genes, which is >3 times the average quantity found in other EPVs (7.6 gene copies ± 4.8 SD) (S4 Table). Furthermore, 14 of the DlEPV BRO genes are clustered together in the most central 30 kb of the genome, while 4 additional BRO genes form a secondary cluster within 17 kb of the primary cluster (Fig 1). The uniform strand orientation and relative size of BRO genes in the primary cluster suggests that this region represents a large gene family expansion via tandem duplication events.

Few other homologous virulence genes were identified. Three DlEPV virulence genes could be identified by sequence similarity to virulence genes in other viruses, including a thymidylate kinase (DLEV176), a thymidine kinase (DLEV178), and a F-box protein (DLEV179) (S1 Table). Thymidine and thymidylate kinases are ubiquitous genes involved in DNA biosynthesis of both eukaryotes and prokaryotes, in which they sequentially phosphorylate nucleosides before incorporation into a growing DNA strand [64]. Although poxvirus thymidine and thymidylate

kinases similarly function in viral DNA replication, they are not highly conserved and are not required for replication of all poxviruses [58,65]. Furthermore, thymidine kinase has a proposed virulence role in VACV due to the drop in pathogenicity associated with thymidine kinase-negative viral recombinants [66]. F-box proteins are commonly found in eukaryotic genomes and contribute to the cellular ubiquitination system for protein degradation [67–69]. F-box-like proteins are also highly abundant virulence genes in CPV genomes, where they interact with host cell ubiquitin-proteasome components and lead to the degradation of important immunity-related proteins, such as nuclear factor kappa B (NF-kB) transcription factors [70–74]. Only one F-box domain-containing gene has been previously reported in an EPV, which is AMV254, a tryptophan repeat gene family protein in the AMEV genome [39]. The DlEPV putative F-box gene shows no similarity to AMV254 and differs from CPV F-box-like genes in that it does not contain the characteristic ankyrin repeats of poxvirus F-box-like genes [71]. In addition, the F-box domain of DLEV179 is located at the N-terminus, which is more similar to eukaryotic F-box proteins [75], rather than the C-terminal location common to viral F-box-like domains [71].

The DlEPV genome also contains three genes with protein domains not found in other viruses that may also be involved in virulence, such as a gamma-glutamyl transpeptidase domain (DLEV037), a type IV secretion system domain (DLEV099), and a thermostable hemolysin domain (DLEV172). The gamma-glutamyl transpeptidase (GGT) gene was first identified in DlEPV by Hashimoto and Lawrence [76]. A GGT has also been identified in the venom of *Aphidius ervi* parasitoid wasps, where it was shown to cause ovarian cell apoptosis in the aphid hosts of the wasp [77]. The DlEPV genes that encode the type IV secretion system (T4SS) and hemolysin domains could also be involved in host cell death, as these domains are used by many bacterial pathogens for cell membrane pore formation. T4SSs are used broadly by bacteria for the transfer of macromolecules, like DNA or proteins, to bacterial or eukaryotic cells [78,79]. Furthermore, hemolysins are toxins specifically used by bacterial pathogens, such as *Vibrio* species, to rupture blood cells of the infected host [80].

Additional putative virulence genes characterize the DlEPV genome. Due to the overall lack of poxvirus virulence gene conservation, we conducted a promoter sequence analysis to identify additional putative virulence genes in the DlEPV genome. Transcription of poxvirus genes is temporally regulated during infection, based on promoter sequence recognition by viral transcription factors that are specific to different stages of the replication cycle [56]. Viral genes transcribed soon after infection are known as early genes and include the virulence genes, which are expressed quickly to combat host defenses against infection [56,81,82]. CPV early genes contain a conserved upstream promoter sequence that is recognized by the early transcription factor packaged within virions [83–85]. Similarly, the promoter sequence TGAAAXXXXA is conserved among EPV early genes [39,41], so we searched the 100 bp upstream region of DlEPV ORFs without an assigned putative function for the EPV early promoter motif. We identified 34 additional putative virulence genes from this approach (Fig 1 and S1 Table). When combined with the 27 BRO genes and 6 virulence genes described above, a total of 67 putative virulence genes were identified in the DlEPV genome.

Functional genomic analysis of DlEPV

We next used our annotation of the DlEPV genome to investigate the differential functionality of this virus in its two insect hosts. We have previously shown that DlEPV replicates in both wasp and fly tissue, but only flies are susceptible to the virulent effects of the virus [34]. In order to maintain a stable symbiosis, we hypothesized that DlEPV activity is altered within the wasp venom gland, such that replication is maximized and virulence is minimized. Conversely, we predicted that DlEPV replication and virulence activity follows a more standard poxvirus

trajectory within the fly host. Because DlEPV shows no evidence of endogenization within the *D. longicaudata* genome, this variation in replication and virulence functions within the wasp would have to be achieved through a different mechanism than the genomic integration and reorganization feature of parasitoid EVEs. We therefore utilized RNA sequencing (RNA-seq) to examine whether differential viral gene expression is associated with the selective virulence demonstrated by DlEPV.

Identification of peak DlEPV expression for transcriptome sequencing. Stages of peak viral gene expression were first determined in both wasp and fly tissues to select for transcriptome sequencing. Our previously reported DlEPV expression profiles from wasp venom gland tissue indicated that viral gene expression peaked in female wasps at the time of the final molt into adulthood [34] (Fig 4A-4C). Here, we generated similar reverse transcription qPCR (RT-qPCR) profiles of whole flies throughout parasitism using the same 3 genes that were analyzed in venom gland tissue: the 147 kDa RNA polymerase subunit RPO147 (DLEV067), the DNA polymerase DNAP (DLEV168), and the P4b structural component (DLEV147). DlEPV gene expression rose steadily in flies 4–24 hours post parasitism (hpp) and plateaued at 48–96 hpp (RPO147 $F_{5,30} = 12.29$, $p < 0.0001$; DNAP $F_{5,30} = 9.82$, $p < 0.0001$; P4b $F_{5,30} = 111.54$, $p < 0.0001$) (Fig 4D-4F), which is congruent with prior qPCR quantification of viral genome copy growth in parasitized flies [34]. Using these data, we determined that the unemerged adult female wasp and the 72 hpp fly were comparable stages of maximum DlEPV expression activity.
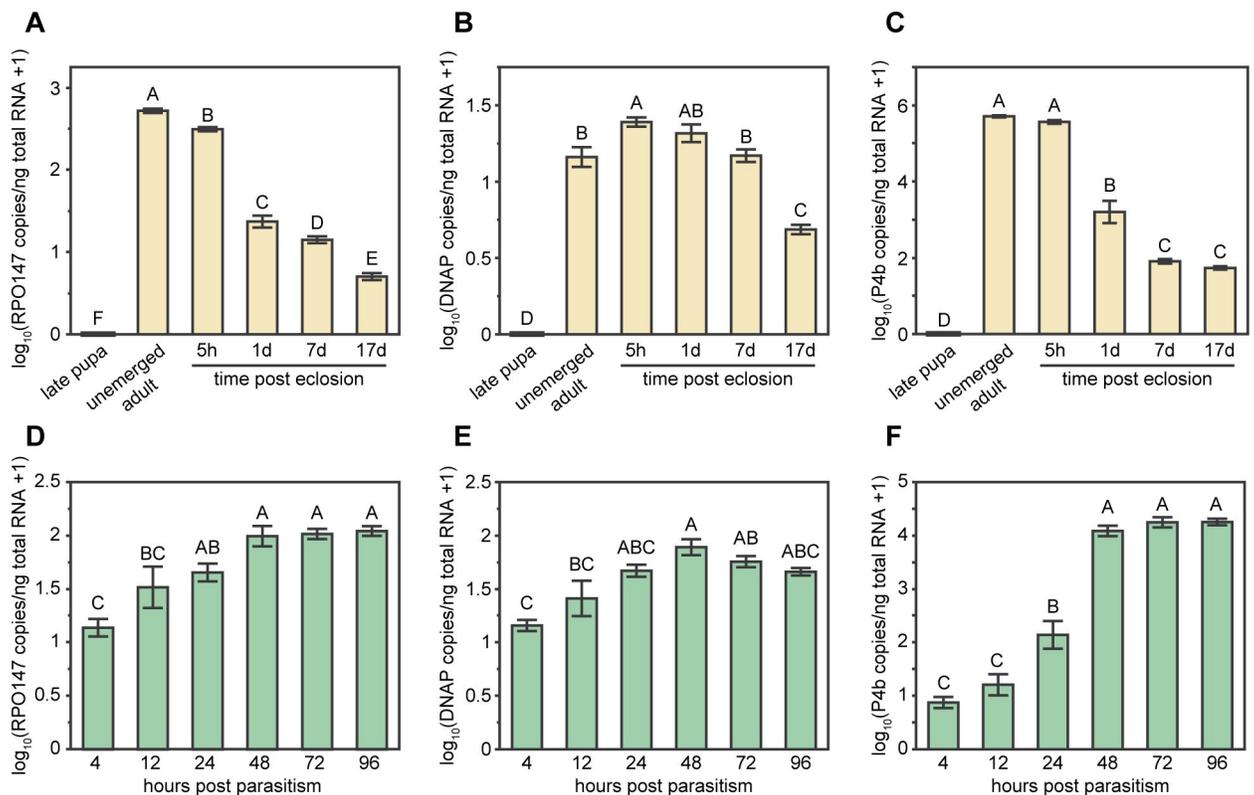
Fig 4. DlEPV gene expression profiles for selection of RNA-seq timepoints. Expression of DlEPV genes measured with RT-qPCR in (A-C) female wasp venom glands from late pupa to 17-day-old adult and (D-F) parasitized flies from 4–96 hours post parasitism (hpp). Profiled DlEPV genes encode (A,D) the 147kDa RNA polymerase subunit RPO147, (B,E) the DNA polymerase DNAP, and (C,F) the structural protein P4b. Each bar represents the mean $\log_{10}$ transformed cDNA copy number per ng total RNA averaged from 6 biological replicates. Error bars represent one standard error above and below the mean, and the letter(s) above each bar indicates statistically distinct mean values from Tukey's HSD tests. Data in (A-C) were modified from Coffman *et al.* 2020 [34].

Differential viral gene expression supports two DlEPV functional roles. Total RNA was isolated from venom gland tissue of unemerged female wasps, as well as whole fly body tissue at 72 hpp (with wasp larva removed), including 6 biological replicates of each treatment. Because we had previously generated a transcriptome from unemerged wasp venom gland tissue [34], only 5 additional replicates were sequenced for this treatment. Paired-end sequencing followed by read quality filtering yielded an average 11.8 million read pairs per wasp sample and 8.9 million read pairs per fly sample. An average 38.9% of venom gland reads and 7.8% of parasitized fly reads aligned to the DlEPV genome. 91.2% (176 of 193) of DlEPV genes showed significant differential expression (FDR, $q < 0.05$) during virus replication in wasps and flies (S5 Table). Hierarchical clustering yielded two main groups of differentially expressed DlEPV genes: 86 genes were significantly upregulated, and 90 genes were significantly downregulated during virus replication in the wasp venom gland compared to the parasitized fly (Fig 5). Genes upregulated in wasp tissue displayed an average $\log_2$ fold change of 2.3, which is nearly a 5x greater level of expression in wasps compared to flies. Even more drastic were genes downregulated in wasp tissue, which had an average $\log_2$ fold change of 3.4, or >10x lower expression in wasps compared to flies (S5 Table). We then looked for differential expression patterns associated with the two main functional gene groups identified in the DLEPV genome: core replication genes and virulence genes. Remarkably, 82.2% (37 of 45) DlEPV core replication genes fell within the former cluster of genes upregulated in wasp tissue. The latter cluster of DlEPV genes downregulated in wasp tissue contained 79.1% (53 of 67) of DlEPV putative virulence genes, including 22 of the 27 BRO genes, 5 of the 6 virulence genes identified by sequence similarity, as well as 26 of the 34 additional virulence genes identified by their early promoter motif.

Discussion

The genomic architecture of PDVs underpins many aspects of their associations with parasitoid wasps, including Mendelian inheritance of the proviral genome, the shift of virus replication control to the wasp, and restriction of replication to wasps and virulence to hosts [27]. We have previously shown that DlEPV broadly shares these features with PDVs as products of convergent evolution [34]. However, our work in this study has shown that DlEPV lacks the fundamental integration event that facilitated the evolution of these characteristics in PDV and other EVE systems. Our findings therefore raise intriguing questions regarding how features like vertical transmission, controlled virus replication, and selective virulence arose and are maintained in the DlEPV system. Furthermore, functional data presented here demonstrate that one of these features, partitioning of viral activity, is accomplished by DlEPV through a method not before observed for a beneficial virus.

DlEPV represents an exogenous parasitoid virus

The presence of wasp genes surrounding multiple viral gene clusters is repeatedly observed with parasitoid EVE genome sequences [17–22]. Conversely, our assembly of the DlEPV genome into a single contig without bordering wasp DNA provides evidence that DlEPV is not integrated into the *D. longicaudata* genome and thus, is not an EVE. qPCR measurements of DlEPV genome copy number normalized by *D. longicaudata* genome copy number provide further support for the non-endogenous status of this virus, because they reveal that DlEPV is not consistently present in all wasp cells, which we would expect if DlEPV was an endogenous provirus. The notion that DlEPV is not integrated into the *D. longicaudata* genome is perhaps not surprising given the atypical replication strategy of other poxviruses. The family *Poxviridae* is unique among most other DNA viruses, as the poxvirus replication cycle is completed entirely within the cytoplasm of infected cells and does not require localization to the nucleus
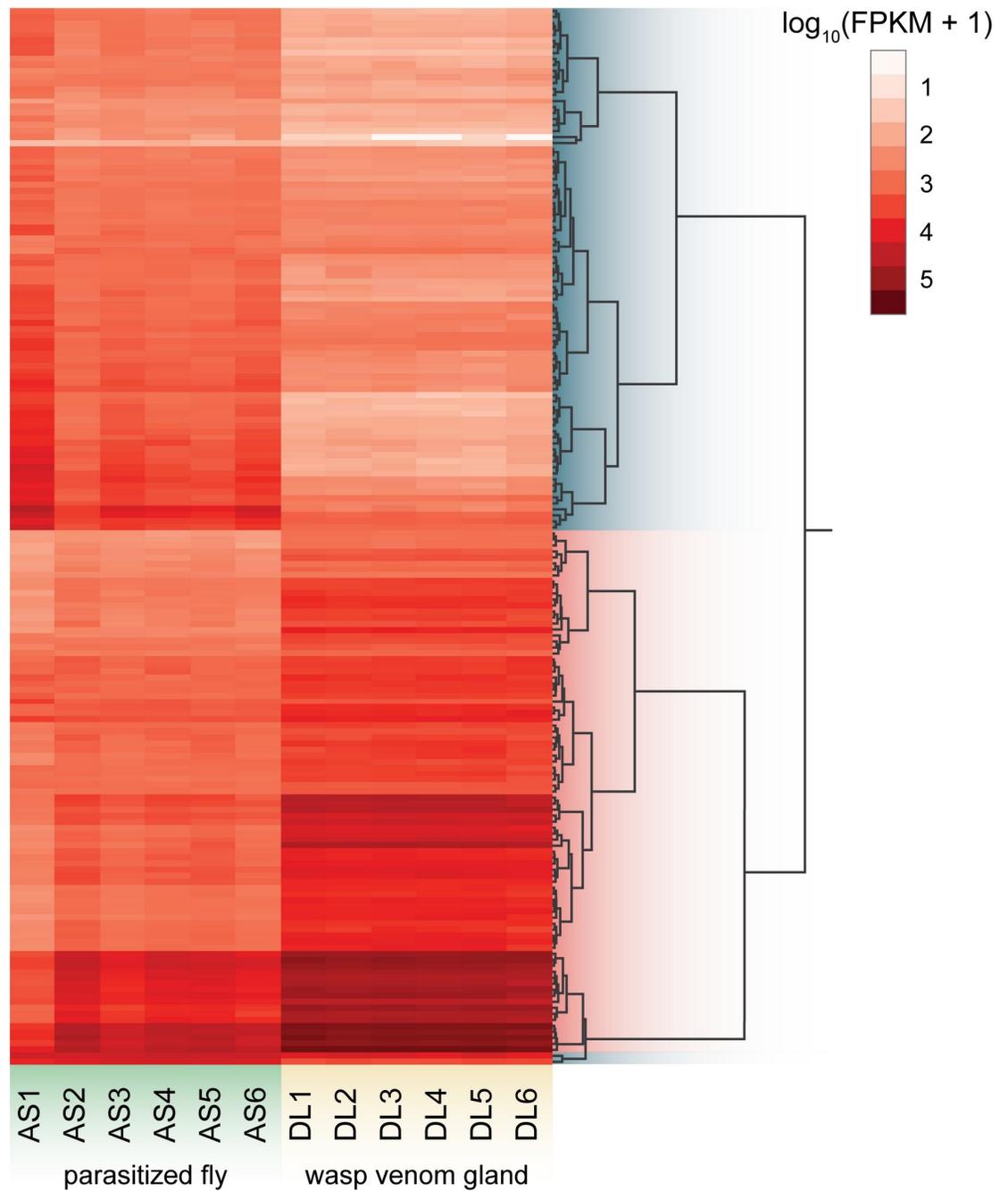
Fig 5. DlEPV shows widespread differential expression during replication in *D. longicaudata* wasps compared to *A. suspensa* flies. Heatmap showing significantly (FDR $q < 0.05$) differentially expressed DlEPV genes in wasp and fly tissue. Each row represents a DlEPV gene, and each column represents that gene's expression in each of the 12 RNA samples. Expression is depicted as the $\log_{10}$ transformed FPKM value. Columns AS1-6 correspond to the 6 parasitized fly RNA replicates, and DL1-6 correspond to the 6 wasp venom gland replicates. Rows were clustered using the Ward method based on similarity in gene expression pattern across the 12 samples: DlEPV genes that were significantly downregulated in wasp tissue are highlighted in blue, and genes that were significantly upregulated in wasp tissue are highlighted in pink.

or integration of viral DNA into the host genome [38]. Nevertheless, the exogenous nature of the DlEPV genome contrasts starkly with the integrated and dispersed viral genome architecture of PDVs and other parasitoid EVEs [14,22].

Our DlEPV genome assembly is similar in total length to other poxvirus genomes, and our annotation of the DlEPV genome yielded the majority of conserved poxvirus core genes, indicating that we have successfully obtained the entire viral genome sequence. However, 4 of the 49 poxvirus core genes were not identified through sequence similarity searches. Given the absence of these genes in the *D. longicaudata* venom gland transcriptome, it is unlikely that these genes have integrated into the wasp genome. In addition, the high sequence divergence of identifiable DlEPV core genes displayed in our phylogeny suggests that the missing core genes still reside within the DlEPV genome but have diverged in sequence past the point of detection by our search methods. Furthermore, the similar level of sequence divergence in Yalta virus core genes, combined with the mutual absence of 2 core genes between the DlEPV and Yalta virus genomes support a lineage-specific divergence of these core genes. We can not rule out, however, that these genes may have instead been lost entirely from this poxvirus lineage and are not required for successful infection and replication within their respective hosts.

DlEPV may have originated as a fly pathogen

Several EPVs with dipteran hosts have been described, but most representatives have been isolated from chironomid midges or mosquitoes and lack genetic sequence data [35]. Yalta virus is the first dipteran poxvirus isolated from the higher flies (suborder Brachycera), which also contains the tephritid fruit flies that serve as hosts for *D. longicaudata* wasps. The close relationship found between DlEPV and Yalta virus in this investigation supports the hypothesis that DlEPV arose from a fly pathogen. However, more taxonomic sampling of fly poxvirus genomes would be required to rule out the possibility that Yalta virus is instead a remnant parasitoid virus within a drosophilid host. How the DlEPV progenitor was acquired by the *D. longicaudata* lineage is unknown but could have occurred through a variety of events, since parasitoid wasps can come into contact with the pathogens of their hosts during development, as well as adulthood. For example, ascoviruses are pathogenic insect DNA viruses exclusively vectored to new lepidopteran hosts via contamination of adult parasitoid wasp ovipositors that are used to lay eggs within them [86]. The precise origin of most other parasitoid viruses and EVEs remains somewhat obscure due to limited taxonomic sampling of closely related insect DNA viruses, but many are suspected to be derived from pathogens of the parasitoids' host insects [87]. The recent discoveries of hymenopteran and *Drosophila* poxviruses have allowed us to conduct a closer examination that suggests DlEPV originated as a pathogen from a host fly of the *D. longicaudata* ancestor. DlEPV thus provides more evidence for how viruses can be acquired by parasitoid wasps and lead to symbiogenesis events.

The genomic differences of DlEPV and Yalta virus compared to other sequenced EPVs suggest that insect poxviruses are more diverse than originally understood. The proximity of DlEPV and Yalta virus to CPVs in our unrooted phylogeny (Fig 2) makes the once clear divide between the EPV and CPV subfamilies appear more ambiguous. Other genomic features shared by DlEPV and Yalta virus differ from EPVs, such as their nucleotide base composition and gene content. Both DlEPV and Yalta virus contain higher than average GC content compared to other EPVs and therefore are more similar to CPVs, which can range widely in GC content from 25–65% [45]. The approximate 40% GC estimated for the *D. longicaudata* genome could also contribute to the elevated GC composition of DlEPV, in particular, due to the vertical transmission of this virus within an insect host to which other EPVs are not subjected. Additionally, apart from the core genes shared by all poxviruses, DlEPV and Yalta virus contain very few of the additional 50 genes shared by all EPVs [40,42]. One notably absent EPV-specific core gene from both DlEPV and Yalta virus genomes is that which encodes the protein spheroidin. This protein is not found in CPVs but is the main component of the

characteristic EPV spheroid occlusion bodies, which are thought to protect EPV virions from environmental inactivation agents, such as UV light [88]. While the spheroidin gene may be divergent to a point beyond detection, a second scenario is that spheroidin is truly absent and not required for successful transmission of these viruses.

Transcriptomic data support DlEPV functional dichotomy and genomic adaptations

The ability of DlEPV to replicate within both wasps and flies but only cause pathogenic effects during fly replication implies that DlEPV virulence is mitigated during virus replication in wasps. This strategy would promote a more stable association between virus and wasp, but as DlEPV is not endogenous, the viral genome integration and dispersion observed in other parasitoid viruses fails to explain how DlEPV completes nonpathogenic replication within wasp tissue. We looked at differential viral gene expression during replication in wasps and flies as an alternative mechanism that might corroborate the selective virulence phenotype of DlEPV. Our findings demonstrate that DlEPV transcriptional activity varies largely during replication in wasps compared to flies, supporting a promotion of virus replication and inhibition of virulence in wasp tissue. These distinct expression patterns are correlated to the different predicted roles of the virus in its two hosts: maximum virus replication in wasp tissue produces an abundance of virions for injection into hosts during oviposition, and restriction of virulence to fly tissue manipulates the host physiology for successful parasitism by the developing wasp. Of note, putative virulence genes, such as the BRO genes, were expressed at extremely low levels compared to other DlEPV genes during virus replication in the venom gland, and they represented many of the most differentially expressed viral genes in wasps compared to flies. We hypothesize that regulatory mechanisms exist within *D. longicaudata* that suppress BRO and other virulence gene expression during virus replication in the venom gland to deter viral pathogenicity. Possible measures of DlEPV control such as this warrant future study, because they differ from what is observed in PDV systems. The differential DlEPV gene expression reported here thus represents convergent evolution with endogenous parasitoid viruses to maintain a separation of viral function that aligns with parasitoid wasp survival and fitness.

Results from our transcriptomes also hint that DlEPV genomic features, such as the BRO gene expansion, are adaptive to symbiotic life. The DlEPV BRO genes are far more extensive in copy number than observed in other EPVs, likely due to a large tandem duplication in the DlEPV genome center. In addition, the majority of DlEPV BRO genes demonstrated upregulation in fly tissue, supporting their involvement in virulence within the fly hosts of *D. longicaudata* wasps. Poxviruses have experimentally been shown to undergo rapid, tandem virulence gene duplications as a means of adaptation [89,90]. Therefore, tandem duplication of the BRO genes may be adaptive to the success of DlEPV, or by extension, the success of the developing wasp that is also fighting for survival within the fly host. Similar to DlEPV, PDV genomes contain large gene families with members that function as host virulence factors, such as the protein tyrosine phosphatases (PTPs) [91]. Several PTP members arose by tandem duplication events, and some also show evidence for positive selection [92]. The gene duplications in DlEPV and PDV genomes may therefore represent similar adaptations to respective hosts due to the shared selective pressures accompanied by their associations with parasitoid wasps.

DlEPV is a true mutualistic viral symbiont of parasitoid wasps

Heritable associations between insects and beneficial microbes are often highly complex, due to the unconventional evolutionary forces that act on host-associated microbes [93]. Vertically transmitted bacterial symbionts that are completely restricted to live within insect hosts often

exhibit extreme genome degradation compared to their free-living relatives [94,95]. Even though this genomic decay causes the bacteria to become dependent on their host for survival, the symbionts still maintain an exogenous genome that replicates and evolves separately from the host genome [94,95]. PDVs are similar to many of these bacteria in that they provide essential functions for their parasitoid wasp hosts. However, PDVs are less commonly considered to be true symbionts, because they do not contain a replicative genome external to the wasp genome [96]. Until now, genetic characterization of heritable parasitoid viruses has challenged the very notion of a 'viral symbiont' given the shared endogenous nature of those currently described. Furthermore, known examples of heritable viruses that are not endogenous, such as DpTV and DcPV, remain to be definitively demonstrated as mutualistic. DlEPV is thus an unprecedented example of a virus that fully meets the requirements of a heritable mutualistic symbiont, including an exogenous genome and a beneficial function within *D. longicaudata* wasps. As the first genuine mutualistic viral symbiont of parasitoid wasps to be characterized, DlEPV shows promise as a tractable system from which to gain valuable knowledge on the viral side of microbial mutualism in insects.

Materials and methods

Viral genome sequencing and assembly

D. longicaudata wasps and *A. suspensa* flies were reared as previously described [97]. Dissected venom glands from six *D. longicaudata* adult female wasps were pooled into one sample to enrich for DlEPV DNA, followed by phenol:chloroform DNA extraction as reported previously [34]. The resulting DNA was subjected to both Pacific BioSciences (PacBio) and Illumina technologies to sequence the DlEPV genome. 7.5 μg of venom gland DNA was used to make a 10 kb insert size library using PacBio standard SMRT library construction chemistry. The PacBio library was sequenced using a 120 min movie on one SMRT cell. PacBio data were analyzed using the smrtanalysis-2.2.0 Amazon Machine Image hosted on Amazon Web Services. 150,283 PacBio reads were filtered to retain 78,011 reads with a minimum read score of 0.8 and length of 100 bp. 615 long reads (>6 kb in length) were pre-assembled and error-corrected by aligning short reads (>500 bp) to the longer reads and taking the consensus with HGAP v3.0 [98]. Following this, 522 long error-corrected reads with an N50 of 9,071 bp were assembled into 19 unitigs to form a draft assembly using the Celera Assembler [99]. All reads were aligned to the assembly to give coverage reports and perform polishing with Quiver from SMRT Analysis v2.2. Each unitig was split into 3 kb pieces and analyzed with blastx against the NCBI non-redundant (nr) protein sequence database (downloaded in September 2014). 7 unitigs of putative EPV origin were selected based upon BLAST results and unitig depth of sequence coverage. These unitigs were compared to each other using blastn, which revealed sets that were almost identical and completely nested within each other, and may have been split apart during assembly due to differing numbers of short repeat sequences in each assembled unitig. The nested unitigs were excluded to retain the longest version of each sequence, resulting in three final unitigs. These unitigs contained areas of overlap ranging from 5–10 kb in length, which were assembled manually to form the full DlEPV genome.

Illumina-compatible library construction was performed using 1 μg of venom gland DNA and the standard protocol with Kapa Biosystems DNA library preparation chemistry. The library was sequenced with 8.6 million 75 bp paired-end reads on an Illumina MiSeq instrument at the Georgia Genomics and Bioinformatics Core (GGBC). Reads were filtered with the fastx toolkit (hannonlab.cshl.edu/fastx_toolkit/) to retain reads with >90% of bases with a PHRED score of 30 or higher for both reads in a pair. 6,243,436 read pairs were mapped against the DlEPV genome assembly with bowtie2 v2.2.4 [100] to correct any potential errors

that arose from PacBio sequencing. Variant SNPs, insertions, and deletions present in the Illumina short read alignment were identified with SAMtools v1.0 [101], as well as through manual inspection. 140 total corrections were made to the reference genome.

DlEPV genome annotation

ORFs with a methionine start codon and a length of at least 50 amino acids were predicted using a combination of Prokka v1.6 [102] and Artemis v16 [103]. ORFs with highly repetitive amino acid sequences were manually discarded as illegitimate proteins. The remaining ORFs were subjected to blastp protein searches against the nr database (downloaded in September 2019), as well as custom BLAST databases composed of poxvirus proteins (taxid: 10240) or strictly EPV proteins (taxid: 10284). Conserved protein domain searches were also conducted against the Pfam database using hmmsearch from HMMER v3.1b1 [104]. An *e*-value cutoff of 0.001 for blastp searches and 0.01 for Pfam searches were used for the bulk of viral gene annotations. These combined searches provided putative functions for 88 of the 193 identified DlEPV ORFs, including 44 of the 49 poxvirus core genes.

To find distant homologs to the poxvirus core genes missed by initial blastp and Pfam searches, we looked for possible matches to the core genes that were beyond our original *e*-value cutoffs but found no additional hits despite these relaxed blastp and Pfam search parameters. We then scoured the DlEPV gene set using core gene hidden Markov model (HMM) searches. HMMs were constructed for each poxvirus core gene by aligning amino acid sequences from available EPV and VACV orthologs with MAFFT v7.215 using the—auto alignment setting [105] and hmmbuild from HMMER. Each HMM was then queried against all DlEPV protein sequences with hmmscan. Only one additional core gene, the L5R entry/ fusion membrane protein (DLEV060), was identified from this approach. We also queried our HMMs against intergenic regions of the DlEPV genome to identify core genes that may have been pseudogenized but found no hits from this approach.

A *de novo* transcriptome assembly was generated from previously published venom gland RNA-seq reads (accession GSE144541) to check if the four remaining core genes had integrated into the *D. longicaudata* genome. First, bowtie2 v2.2.4 was used to map quality filtered reads from the venom gland transcriptome to the DlEPV genome. Reads that failed to map to the reference genome were collected and fed as input for *de novo* transcript assembly using Trinity v2.0.6 [106]. A BLAST nucleotide database was created from the resulting assembly, and the missing core genes were queried against it with tblastn. We then queried the HMMs of the missing genes against the translated transcriptome assembly with HMMER but found no significant hits from either approach.

Comparative genomic analyses

We used publicly available annotations to calculate genome metrics for the majority of poxvirus genomes featured in this study: ACEV (accession NC_023426), AHEV (NC_021247), AMEV (NC_002520), CBEV (NC_021248), CREV (NC_021249), MSEV (NC_001993), MySEV (NC_021246), Yalta virus (MT364305), VACV (NC_006998), ORFV (NC_005336), MOCV (NC_001731), FWPV (NC_002188), CRV (NC_008030), and SGPV (NC_027707). However, the partial LHEV genome (NC_040577) required re-annotation for use in our analyses. We used Prokka v1.13 to call ORFs within the 46 kb LHEV genome segment and identified 53 total ORFs (S6 Table). BLAST searches against poxvirus and EPV protein databases yielded 18 LHEV ORFs that showed similarity to poxvirus core genes. There were two instances in which two adjacent ORFs had sequence similarity to opposite ends of the same core gene. In both cases, the two ORFs had the same strand orientation and were separated by a single

frame shift. Therefore, we assumed the original core gene was incorrectly split into two ORFs due to a single nucleotide sequence error. A total of 16 unique core genes were thus identified from the partial LHEV genome.

Nucleotide composition (% GC) was estimated for the *D. longicaudata* and *A. suspensa* genomes using RNA-seq transcriptomes that were generated as described below. For each species, Trinity was used to construct a *de novo* assembly from RNA reads that failed to map to the DlEPV genome combined for all 6 RNA replicate samples per insect. GC content was then calculated from the resulting assemblies.

To generate the poxvirus core gene phylogeny, amino acid sequences for EPV and CPV orthologs of the 16 core genes found in LHEV (S3 Table) were aligned with MAFFT—auto, concatenated using Geneious Prime 2019.2.3 (https://www.geneious.com), and trimmed of alignment positions in which >50% of taxa contained a gap using trimAl v1.4.1 [107]. The ML phylogenetic tree was generated using RAxML v8.2.11 [108], in which the Gamma model of rate heterogeneity and the LG amino acid substitution matrix were utilized.

RNA isolation and RT-qPCR estimation of viral gene expression

DlEPV gene expression in host flies during parasitism was measured by offering third instar fly larvae to 7-day-old adult wasps that had no prior oviposition experience for 2 h. Resulting flies containing a single laid wasp egg (i.e. those with one oviposition scar) were collected at 4–96 hpp. Flies were kept in standard rearing conditions until each sampling time point. Whole fly samples were collected in a guanidine hydrochloride lysis buffer consisting of 4.9M guanidine hydrochloride, 2% sarkosyl, 50 mM Tris-Cl (pH 7.6), and 10 mM EDTA. Total RNA was isolated using phenol:chloroform, followed by DNase treatment with the TURBO DNA-*free* Kit (Ambion), and elution in 30 μL water. First-strand cDNA was synthesized with 1,000 ng fly RNA according to the Superscript III reverse transcriptase protocol (Invitrogen) using oligo(dT) primers. qPCR reactions were performed as described previously for wasp venom gland DlEPV expression profiling [34]. JMP Pro 14 was used for statistical analysis of RT-qPCR data. One-way ANOVA was used to test for differences in means of biological replicates, and Tukey's HSD was used for multiple comparison tests. Copy numbers were $\log_{10}$ transformed prior to statistical analysis.

Transcriptome sequencing and analysis

Unemerged wasp venom glands were collected in triplicate as was done for our initial venom gland transcriptome to obtain 5 additional venom gland samples for a total of 6 biological replicates. Singly-scarred fly pupae were collected 72 hpp by first removing the developing wasp larva by dissection in PBS. A total of 6 flies were collected in this manner with each specimen representing one biological replicate. Total RNA for sequencing was extracted using the RNeasy Mini Kit (QIAGEN) with on-column DNase digestion, followed by a secondary DNase treatment using the TURBO DNA-*free* Kit (Ambion) after RNA isolation. Illumina-compatible stranded RNA libraries for the 12 samples were constructed at the GGBC with the Kapa Biosystems RNA library preparation chemistry using 3 μg RNA from each fly sample and 1μg RNA from each venom gland sample. Libraries were sequenced using an Illumina NextSeq instrument at the GGBC, which generated an average of 18.5 million and 14.2 million 75 bp paired-end reads for each wasp and fly sample, respectively. Reads were quality filtered using the fastx toolkit as described before with the Illumina DNA sequencing. Quality reads for each sample were separately mapped to the DlEPV reference genome using bowtie2 v2.2.4. Cuffquant from Cufflinks v2.2.1 was used to calculate the average fragments per kilobase of transcript per million mapped reads (FPKM) values for each DlEPV ORF in both wasp and fly

tissues, and Cuffdiff was used to test for differential expression between the two treatments [109]. Differential expression for a gene was considered significant for FDR-adjusted q-values < 0.05 [110]. Hierarchical clustering of the significantly differentially expressed DlEPV genes was performed with JMP Pro 14.

Supporting information

S1 Table. Annotated ORFs in the DlEPV genome.
(PDF)

S2 Table. Sequenced EPV genome features.
(PDF)

S3 Table. Poxvirus core gene homologs in EPV and CPV genomes. The 49 poxvirus core genes are shown with corresponding locus tags for homologous ORFs in EPV and CPV genomes. The 16 core genes used to build the phylogeny in Fig 2 are highlighted in yellow.
(PDF)

S4 Table. BRO genes in EPV genomes. BRO genes are defined as those with a Bro-N protein domain. Protein domains were identified using hmmsearch to query genes from each genome against the Pfam database. A maximum e-value cutoff of 0.05 was used to isolate significant domain matches.
(PDF)

S5 Table. Expression of DlEPV genes in wasp venom gland (DL) and parasitized fly (AS) samples. Genes are grouped by their putative function based on similarity to other poxvirus genes. Genes of putative virulence function are subdivided between those with a Bro-N domain (Virulence: BRO Genes), those that were identified by sequence similarity to known virulence genes (Virulence: Homology), and those that have a conserved EPV early gene promoter motif and no other assigned function (Virulence: Early Promoter). Genes with an asterisk indicate those that demonstrated significant differential expression between the two treatments ($q < 0.05$).
(PDF)

S6 Table. Re-annotation of the LHEV genome segment. Feature table of LHEV ORFs including the 11 ORFs previously annotated by Viljakainen *et al.* 2018 [43] (accession NC_040577).
(PDF)

S1 Fig. Normalized DlEPV abundance in *D. longicaudata* wasps. DlEPV copy number relative to *D. longicaudata* copy number was estimated with qPCR for (A) adult female wasp reproductive tissues, and (B) female and male whole wasps in pupal-adult developmental stages. Venom glands and ovaries from adult females were pooled in triplicate for each biological replicate. DlEPV genome copy number was estimated using the poly(A) polymerase small subunit gene (polyAPol, DLEV167), and *D. longicaudata* copy number with the elongation factor alpha gene (EF1a). qPCR was performed as done previously [34]. The y-axes indicate the $\log_{10}$ fold change of total DlEPV genome copy number over total *D. longicaudata* genome copy number. Permanent integration of the DlEPV genome into the *D. longicaudata* genome would result in a ratio of virus to wasp copy number that is ≥ 1 for all wasp tissues, developmental stages, and sexes, which is equivalent to a $\log_{10}$ abundance fold change of 0. Negative $\log_{10}$ abundance fold change values indicate samples in which the virus to wasp copy number ratio was < 1. Each bar represents the average relative DlEPV copy number across 6 biological

replicates, and error bars represent one standard error above and below the mean.
(PDF)

S2 Fig. Additional poxvirus core gene phylogenies. (A) Maximum likelihood (ML) phylogenetic tree built from a concatenated multiple sequence alignment of the 44 core genes shared by all EPV complete genomes. Methods were the same as used to build the phylogeny in Fig 2. Node support (%) was inferred with 1,000 bootstrap iterations. (B) Bayesian inference phylogeny of the concatenated 16 core genes from Fig 2 built using PhyloBayes-MPI v20161021 with the CAT-GTR substitution model [111]. Node support in panel B is labeled with the consensus posterior probability from two independent Markov chain Monte Carlo simulations that ran for 10,000 cycles each with a 1,000 cycle burn-in. (C) ML phylogeny built from a concatenated alignment of the 10 poxvirus core genes shared by all poxviruses and sister group Asfarviridae members African swine fever virus (ASFV, NC_001659) and Kaumoebavirus (NC_034249). Core genes and respective ASFV and Kaumoebavirus accession numbers used to create the phylogeny in panel C include the DNA polymerase E9L (AAA65319.1, ARA71927.1), RNA helicase I8R (AAA65302.1, ARA71975.1), RPO147 J6R (AAA65328.1, ARA71945.1 and ARA71948.1), mRNA-capping enzyme large subunit D1R (AAA65330.1, ARA71993.1), NTPase, DNA primase D5R (AAA65301.1, ARA71965.1), viral early transcription factor (VETF) small subunit D6R (AAA65335.1, ARA72203.1), VETF large subunit A7L (AAA65318.1, ARA71923.1), ATPase NPH1 D11R (AAA65350.1, ARA72259.1), RPO132 A24R (AAA65283.1, ARA72182.1), and ATPase/DNA-packaging protein A32L (AAA65308.1, ARA72015.1). Tree building methods were the same as done for other ML trees.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Kelsey A. Coffman, Gaelen R. Burke.

Data curation: Kelsey A. Coffman, Gaelen R. Burke.

Formal analysis: Kelsey A. Coffman, Gaelen R. Burke.

Funding acquisition: Gaelen R. Burke.

Investigation: Kelsey A. Coffman, Gaelen R. Burke.

Methodology: Kelsey A. Coffman, Gaelen R. Burke.

Project administration: Kelsey A. Coffman, Gaelen R. Burke.

Resources: Kelsey A. Coffman, Gaelen R. Burke.

Software: Kelsey A. Coffman, Gaelen R. Burke.

Supervision: Kelsey A. Coffman, Gaelen R. Burke.

Validation: Kelsey A. Coffman, Gaelen R. Burke.

Visualization: Kelsey A. Coffman, Gaelen R. Burke.

Writing – original draft: Kelsey A. Coffman.

Writing – review & editing: Kelsey A. Coffman, Gaelen R. Burke.

References

1. Moran NA. Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci U S A. 2007; 104:8627–8633. https://doi.org/10.1073/pnas.0611659104 PMID: 17494762

2. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A. 2013; 110:3229–3236. https://doi.org/10.1073/pnas.1218525110 PMID: 23391737

3. Douglas AE. Multiorganismal insects: diversity and function of resident microorganisms. Annu Rev Entomol. 2015; 60:17–34. https://doi.org/10.1146/annurev-ento-010814-020822 PMID: 25341109

4. Baumann P. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol. 2005; 59:155–189. https://doi.org/10.1146/annurev.micro.59.030804.121041 PMID: 16153167

5. Rio RVM, Attardo GM, Weiss BL. Grandeur alliances: symbiont metabolic integration and obligate arthropod hematophagy. Trends Parasitol. 2016; 32:739–749. https://doi.org/10.1016/j.pt.2016.05.002 PMID: 27236581

6. Engel P, Moran NA. The gut microbiota of insects–diversity in structure and function. FEMS Microbiol Rev. 2013; 37:699–735. https://doi.org/10.1111/1574-6976.12025 PMID: 23692388

7. Roossinck MJ. The good viruses: viral mutualistic symbioses. Nat Rev Microbiol. 2011; 9:99–108. https://doi.org/10.1038/nrmicro2491 PMID: 21200397

8. Gauthier J, Drezen J-M, Herniou EA. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. Parasitology. 2018; 145:713–723. https://doi.org/10.1017/S0031182017000725 PMID: 28534452

9. Dicke M, Cusumano A, Poelman EH. Microbial symbionts of parasitoids. Annu Rev Entomol. 2020; 65:171–190. https://doi.org/10.1146/annurev-ento-011019-024939 PMID: 31589823

10. Strand MR, Pech LL. Immunological basis for compatibility in parasitoid host relationships. Annu Rev Entomol. 1995; 40:31–56. https://doi.org/10.1146/annurev.en.40.010195.000335 PMID: 7810989

11. Pennacchio F, Strand MR. Evolution of developmental strategies in parasitic Hymenoptera. Annu Rev Entomol. 2006; 51:233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029 PMID: 16332211

12. Asgari S, Rivers DB. Venom proteins from endoparasitoid wasps and their role in host-parasite interactions. Annu Rev Entomol. 2011; 56:313–335. https://doi.org/10.1146/annurev-ento-120709-144849 PMID: 20822448

13. Strand MR, Burke GR. Polydnavirus-wasp associations: evolution, genome organization, and function. Curr Opin Virol. 2013; 3:587–594. https://doi.org/10.1016/j.coviro.2013.06.004 PMID: 23816391

14. Drezen J-M, Leobold M, Bézier A, Huguet E, Volkoff A-N, Herniou EA. Endogenous viruses of parasitic wasps: variations on a common theme. Curr Opin Virol. 2017; 25:41–48. https://doi.org/10.1016/j.coviro.2017.07.002 PMID: 28728099

15. Beckage NE. Polydnaviruses as endocrine regulators. In: Beckage NE, Drezen J-M, editors. Parasitoid Viruses. San Diego, CA: Academic Press;2012. pp. 163–168.

16. Strand MR. Polydnavirus gene products that interact with the host immune system. In: Beckage NE, Drezen J-M, editors. Parasitoid Viruses. San Diego, CA: Academic Press; 2012. pp. 149–161.

17. Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, et al. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science. 2009; 323:926–930. https://doi.org/10.1126/science.1166788 PMID: 19213916

18. Volkoff A-N, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, et al. Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. PLoS Pathog. 2010; 6:e1000923. https://doi.org/10.1371/journal.ppat.1000923 PMID: 20523890

19. Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet. 2014; 10:e1004660. https://doi.org/10.1371/journal.pgen.1004660 PMID: 25232843

20. Béliveau C, Cohen A, Stewart D, Periquet G, Djoumad A, Kuhn L, et al. Genomic and proteomic analyses indicate that banchine and campoplegine polydnaviruses have similar, if not identical, viral ancestors. J Virol. 2015; 89:8909–8921. https://doi.org/10.1128/JVI.01001-15 PMID: 26085165

21. Pichon A, Bézier A, Urbach S, Aury JM, Jouan V, Ravallec M, et al. Recurrent DNA virus domestication leading to different parasite virulence strategies. Sci Adv. 2015; 1:e1501150. https://doi.org/10.1126/sciadv.1501150 PMID: 26702449

22. Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. Rapid viral symbiogenesis via changes in parasitoid wasp genome architecture. Mol Biol Evol. 2018; 35:2463–2474. https://doi.org/10.1093/molbev/msy148 PMID: 30053110

23. Burke GR. Common themes in three independently derived endogenous nudivirus elements in parasitoid wasps. Curr Opin Insect Sci. 2019; 32:28–35. https://doi.org/10.1016/j.cois.2018.10.005 PMID: 31113628

24. Strand MR. Polydnaviruses. In: Asgari S, Johnson KN, editors. Insect Virology. Norfolk, UK: Caister Academic Press;2010. pp. 171–197.

25. Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen J-M. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos Trans R Soc Lond B Biol Sci. 2013; 368:20130051. https://doi.org/10.1098/rstb.2013.0051 PMID: 23938758

26. Strand MR, Burke GR. Polydnaviruses as symbionts and gene delivery systems. PLoS Pathog. 2012; 8:e1002757. https://doi.org/10.1371/journal.ppat.1002757 PMID: 22792063

27. Strand MR, Burke GR. Polydnaviruses: nature's genetic engineers. Annu Rev Virol. 2014; 1:333–354. https://doi.org/10.1146/annurev-virology-031413-085451 PMID: 26958725

28. Bigot Y, Rabouille A, Doury G, Sizaret PY, Delbost F, Hamelin MH, et al. Biological and molecular features of the relationships between Diadromus pulchellus ascovirus, a parasitoid hymenopteran wasp (*Diadromus pulchellus*) and its lepidopteran host, *Acrolepiopsis assectella*. J Gen Virol. 1997; 78:1149–1163. https://doi.org/10.1099/0022-1317-78-5-1149 PMID: 9152436

29. Renault S, Petit A, Benedet F, Bigot S, Bigot Y. Effects of the Diadromus pulchellus ascovirus, DpAV-4, on the hemocytic encapsulation response and capsule melanization of the leek-moth pupa, *Acrolepiopsis assectella*. J Insect Physiol. 2002; 48:297–302. https://doi.org/10.1016/s0022-1910(01)00174-3 PMID: 12770103

30. Bigot Y, Renault S, Nicolas J, Moundras C, Demattei MV, Samain S, et al. Symbiotic virus at the evolutionary intersection of three types of large DNA viruses; iridoviruses, ascoviruses, and ichnoviruses. PLoS One. 2009; 4:e6397. https://doi.org/10.1371/journal.pone.0006397 PMID: 19636425

31. Dheilly NM, Maure F, Ravallec M, Galinier R, Doyon J, Duval D, et al. Who is the puppet master? Replication of a parasitic wasp-associated virus correlates with host behaviour manipulation. Proc Biol Sci. 2015; 282:20142773. https://doi.org/10.1098/rspb.2014.2773 PMID: 25673681

32. Edson KM, Barlin MR, Vinson SB. Venom apparatus of braconid wasps: comparative ultrastructure of reservoirs and gland filaments. Toxicon. 1982; 20:553–562. https://doi.org/10.1016/0041-0101(82)90049-6 PMID: 7101306

33. Lawrence PO, Akin D. Virus-like particles from the poison glands of the parasitic wasp *Biosteres longicaudatus* (Hymenoptera: Braconidae). Can J Zool. 1990; 68:539–546.

34. Coffman KA, Harrell TC, Burke GR. A mutualistic poxvirus exhibits convergent evolution with other heritable viruses in parasitoid wasps. J Virol. 2020; 94:e02059–19. https://doi.org/10.1128/JVI.02059-19 PMID: 32024779

35. Skinner MA, Buller RM, Damon IK, Lefkowitz EJ, McFadden G, McInnes CJ, et al. Poxviridae. In: King AMQ, Lefkowitz E, Adams MJ, Carstens EB, editors. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier;2012. pp. 291–309.

36. Smith GL. Genus *Orthopoxvirus*: vaccinia virus. In: Mercer AA, Schmidt A, Weber O, editors. Poxviruses. Basel, CH: Birkhäuser Verlag;2007. pp. 1–45.

37. Perera S, Li Z, Pavlik L, Arif B. Entomopoxviruses. In: Asgari S, Johnson KN, editors. Insect Virology. Norfolk, UK: Caister Academic Press;2010. pp. 83–102.

38. Moss B. Poxviridae: the viruses and their replication. 5th ed. In: Knipe DM, Howley PM, editors. Fields Virology. 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins;2007. pp. 2905–2945.

39. Bawden AL, Glassberg KJ, Diggans J, Shaw R, Farmerie W, Moyer RW. Complete genomic sequence of the Amsacta moorei entomopoxvirus: analysis and comparison with other poxviruses. Virology. 2000; 274:120–139. https://doi.org/10.1006/viro.2000.0449 PMID: 10936094

40. Thézé J, Takatsuka J, Li Z, Gallais J, Doucet D, Arif B, et al. New insights into the evolution of *Entomopoxvirinae* from the complete genome sequences of four entomopoxviruses infecting *Adoxophyes honmai*, *Choristoneura biennis*, *Choristoneura rosaceana*, and *Mythimna separata*. J Virol. 2013; 87:7992–8003. https://doi.org/10.1128/JVI.00453-13 PMID: 23678178

41. Afonso CL, Tulman ER, Lu Z, Oma E, Kutish GF, Rock DL. The genome of Melanoplus sanguinipes entomopoxvirus. J Virol. 1999; 73:533–552. https://doi.org/10.1128/JVI.73.1.533-552.1999 PMID: 9847359

42. Mitsuhashi W, Miyamoto K, Wada S. The complete genome sequence of the Alphaentomopoxvirus Anomala cuprea entomopoxvirus, including its terminal hairpin loop sequences, suggests a potentially unique mode of apoptosis inhibition and mode of DNA replication. Virology. 2014; 452:95–116. https://doi.org/10.1016/j.virol.2013.12.036 PMID: 24606687

43. Viljakainen L, Holmberg I, Abril S, Jurvansuu J. Viruses of invasive Argentine ants from the European Main supercolony: characterization, interactions and evolution. J Gen Virol. 2018; 99:1129–1140. https://doi.org/10.1099/jgv.0.001104 PMID: 29939128

44. Wallace MA, Coffman KA, Gilbert C, Ravindran S, Albery GF, Abbott J, et al. The discovery, distribution and diversity of DNA viruses associated with *Drosophila melanogaster* in Europe. bioRxiv: 2020.10.16.342956v1 [Preprint]. 2020 [cited 2020 Nov 3]. Available from: https://www.biorxiv.org/content/10.1101/2020.10.16.342956v1

45. Lefkowitz EJ, Wang C, Upton C. Poxviruses: past, present and future. Virus Res. 2006; 117:105–118. https://doi.org/10.1016/j.virusres.2006.01.016 PMID: 16503070

46. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol. 2006; 62:551–563. https://doi.org/10.1007/s00239-005-0221-1 PMID: 16557338

47. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. J Virol. 2003; 77:7590–7600. https://doi.org/10.1128/jvi.77.13.7590-7600.2003 PMID: 12805459

48. Gubser C, Hué S, Kellam P, Smith GL. Poxvirus genomes: a phylogenetic analysis. J Gen Virol. 2004; 85:105–117. https://doi.org/10.1099/vir.0.19565-0 PMID: 14718625

49. Goebel SJ, Johnson GP, Perkus ME, Davis SW, Winslow JP, Paoletti E. The complete DNA sequence of vaccinia virus. Virology. 1990; 179:247–266. https://doi.org/10.1016/0042-6822(90)90294-2 PMID: 2219722

50. Lin CL, Chung CS, Heine HG, Chang W. Vaccinia virus envelope H3L protein binds to cell surface heparan sulfate and is important for intracellular mature virion morphogenesis and virus infection in vitro and in vivo. J Virol. 2000; 74:3353–3365. https://doi.org/10.1128/jvi.74.7.3353-3365.2000 PMID: 10708453

51. da Fonseca FG, Wolffe EJ, Weisberg A, Moss B. Characterization of the vaccinia virus H3L envelope protein: topology and posttranslational membrane insertion via the C-terminal hydrophobic tail. J Virol. 2000; 74:7508–7517. https://doi.org/10.1128/jvi.74.16.7508-7517.2000 PMID: 10906204

52. da Fonseca FG, Wolffe EJ, Weisberg A, Moss B. Effects of deletion or stringent repression of the H3L envelope gene on vaccinia virus replication. J Virol. 2000; 74:7518–7528. https://doi.org/10.1128/jvi.74.16.7518-7528.2000 PMID: 10906205

53. Boyd O, Turner PC, Moyer RW, Condit RC, Moussatche N. The E6 protein from vaccinia virus is required for the formation of immature virions. Virology. 2010; 399:201–211. https://doi.org/10.1016/j.virol.2010.01.012 PMID: 20116821

54. Condit RC, Moussatche N. The vaccinia virus E6 protein influences virion protein localization during virus assembly. Virology. 2015; 482:147–156. https://doi.org/10.1016/j.virol.2015.02.056 PMID: 25863879

55. Senkevich TG, Wyatt LS, Weisberg AS, Koonin EV, Moss B. A conserved poxvirus NlpC/P60 superfamily protein contributes to vaccinia virus virulence in mice but not to replication in cell culture. Virology. 2008; 374:506–514. https://doi.org/10.1016/j.virol.2008.01.009 PMID: 18281072

56. Broyles SS. Vaccinia virus transcription. J Gen Virol. 2003; 84:2293–2303. https://doi.org/10.1099/vir.0.18942-0 PMID: 12917449

57. Koonin EV, Yutin N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. Adv Virus Res. 2019; 103:167–202. https://doi.org/10.1016/bs.aivir.2018.09.002 PMID: 30635076

58. McLysaght A, Baldi PF, Gaut BS. Extensive gene gain associated with adaptive evolution of poxviruses. Proc Natl Acad Sci U S A. 2003; 100:15655–15660. https://doi.org/10.1073/pnas.2136653100 PMID: 14660798

59. Nazarian SH, McFadden G. Immunomodulation by poxviruses. In: Mercer AA, Schmidt A, Weber O, editors. Poxviruses. Basel, CH: Birkhäuser Verlag;2007. pp. 273–296.

60. Iyer LA, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 2006; 117:156–184. https://doi.org/10.1016/j.virusres.2006.01.009 PMID: 16494962

61. Bideshi DK, Renault S, Stasiak K, Federici BA, Bigot Y. Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. J Gen Virol. 2003; 84:2531–2544. https://doi.org/10.1099/vir.0.19256-0 PMID: 12917475

62. Kang WY, Suzuki M, Zemskov E, Okano K, Maeda S. Characterization of baculovirus repeated open reading frames (bro) in Bombyx mori nucleopolyhedrovirus. J Virol. 1999; 73:10339–10345. https://doi.org/10.1128/JVI.73.12.10339-10345.1999 PMID: 10559352

63. Zemskov EA, Kang WK, Maeda S. Evidence for nucleic acid binding ability and nucleosome associa-tion of Bombyx mori nucleopolyhedrovirus BRO proteins. J Virol. 2000; 74:6784–6789. https://doi.org/10.1128/jvi.74.15.6784-6789.2000 PMID: 10888617

64. Anderson EP. Nucleoside and nucleotide kinases. In: Boyer PD, editor. The Enzymes. Academic Press;1973. pp. 49–96.

65. Gentry GA. Viral thymidine kinases and their relatives. Pharmacol Ther. 1992; 54:319–355. https://doi.org/10.1016/0163-7258(92)90006-l PMID: 1334563

66. Buller RML, Smith GL, Cremer K, Notkins AL, Moss B. Decreased virulence of recombinant vaccinia virus expression vectors is associated with a thymidine kinase-negative phenotype. Nature. 1985; 317:813–815. https://doi.org/10.1038/317813a0 PMID: 4058585

67. Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, et al. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell. 1996; 86:263–274. https://doi.org/10.1016/s0092-8674(00)80098-7 PMID: 8706131

68. Skowyra D, Craig KL, Tyers M, Elledge SJ, Harper JW. F-box proteins are receptors that recruit phos-phorylated substrates to the SCF ubiquitin-ligase complex. Cell. 1997; 91:209–219. https://doi.org/10.1016/s0092-8674(00)80403-1 PMID: 9346238

69. Kipreos ET, Pagano M. The F-box protein family. Genome Biol. 2000; 1: reviews3002. https://doi.org/10.1186/gb-2000-1-5-reviews3002 PMID: 11178263

70. Camus-Bouclainville C, Fiette L, Bouchiha S, Pignolet A, Counor D, Filipe U, et al. A virulence factor of myxoma virus colocalizes with NF-kappa B in the nucleus and interferes with inflammation. J Virol. 2004; 78:2510–2516. https://doi.org/10.1128/jvi.78.5.2510-2516.2004 PMID: 14963153

71. Mercer AA, Fleming SB, Ueda N. F-box-like domains are present in most poxvirus ankyrin repeat pro-teins. Virus Genes. 2005; 31:127–133. https://doi.org/10.1007/s11262-005-1784-z PMID: 16025237

72. Hsiao JC, Chao CC, Young MJ, Chang YT, Cho EC, Chang W. A poxvirus host range protein, CP77, binds to a cellular protein, HMG20A, and regulates its dissociation from the vaccinia virus genome in CHO-K1 cells. J Virol. 2006; 80:7714–7728. https://doi.org/10.1128/JVI.00207-06 PMID: 16840350

73. Sonnberg S, Seet BT, Pawson T, Fleming SB, Mercer AA. Poxvirus ankyrin repeat proteins are a unique class of F-box proteins that associate with cellular SCF1 ubiquitin ligase complexes. Proc Natl Acad Sci U S A. 2008; 105:10955–10960. https://doi.org/10.1073/pnas.0802042105 PMID: 18667692

74. Mohamed MR, Rahman MM, Lanchbury JS, Shattuck D, Neff C, Dufford M, et al. Proteomic screening of variola virus reveals a unique NF-kappa B inhibitor that is highly conserved among pathogenic orthopoxviruses. Proc Natl Acad Sci U S A. 2009; 106:9045–9050. https://doi.org/10.1073/pnas.0900452106 PMID: 19451633

75. Jin JP, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic analysis and nomen-clature of mammalian F-box proteins. Genes Dev. 2004; 18:2573–2580. https://doi.org/10.1101/gad.1255304 PMID: 15520277

76. Hashimoto Y, Lawrence PO. Comparative analysis of selected genes from Diachasmimorpha longi-caudata entomopoxvirus and other poxviruses. J Insect Physiol. 2005; 51:207–220. https://doi.org/10.1016/j.jinsphys.2004.10.010 PMID: 15749105

77. Falabella P, Riviello L, Caccialupi P, Rossodivita T, Valente MT, De Stradis ML, et al. A gamma-gluta-myl transpeptidase of *Aphidius ervi* venom induces apoptosis in the ovaries of host aphids. Insect Bio-chem Mol Biol. 2007; 37:453–465. https://doi.org/10.1016/j.ibmb.2007.02.005 PMID: 17456440

78. Voth DE, Broederdorf LJ, Graham JG. Bacterial Type IV secretion systems: versatile virulence machines. Future Microbiol. 2012; 7:241–257. https://doi.org/10.2217/fmb.11.150 PMID: 22324993

79. Grohmann E, Christie PJ, Waksman G, Backert S. Type IV secretion in Gram-negative and Gram-pos-itive bacteria. Mol Microbiol. 2018; 107:455–471. https://doi.org/10.1111/mmi.13896 PMID: 29235173

80. Zhang X-H, Austin B. Haemolysins in *Vibrio* species. J Appl Microbiol. 2005; 98:1011–1019. https://doi.org/10.1111/j.1365-2672.2005.02583.x PMID: 15836469

81. Yang ZL, Bruno DP, Martens CA, Porcella SF, Moss B. Simultaneous high-resolution analysis of vac-cinia virus and host cell transcriptomes by deep RNA sequencing. Proc Natl Acad Sci U S A. 2010; 107:11513–11518. https://doi.org/10.1073/pnas.1006594107 PMID: 20534518

82. Yang ZL, Cao S, Martens CA, Porcella SF, Xie Z, Ma M, et al. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. J Virol. 2015; 89:6874–6886. https://doi.org/10.1128/JVI.00528-15 PMID: 25903347

83. Broyles SS, Yuen L, Shuman S, Moss B. Purification of a factor required for transcription of vaccinia virus early genes. J Biol Chem. 1988; 263:10754–10760. PMID: 3392040

84. Broyles SS, Li J, Moss B. Promoter DNA contacts made by the vaccinia virus early transcription factor. J Biol Chem. 1991; 266:15539–15544. PMID: 1869571

85. Davison AJ, Moss B. Structure of vaccinia virus early promoters. J Mol Biol. 1989; 210:749–769. https://doi.org/10.1016/0022-2836(89)90107-1 PMID: 2515286

86. Bideshi DK, Bigot Y, Federici BA, Spears T. Ascoviruses. In: Asgari S, Johnson KN, editors. Insect Virology. Caister Academic Press; 2010. pp. 3–34.

87. Whitfield JB, Asgari S. Virus or not? Phylogenetics of polydnaviruses and their wasp carriers. J Insect Physiol. 2003; 49:397–405. https://doi.org/10.1016/s0022-1910(03)00057-x PMID: 12770619

88. Becker MN, Moyer RW. Subfamily *Entomopoxvirinae*. In: Mercer AA, Schmidt A, Weber O, editors. Poxviruses. Basel, CH: Birkhäuser Verlag; 2007. pp. 253–271.

89. Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, et al. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell. 2012; 150:831–841. https://doi.org/10.1016/j.cell.2012.05.049 PMID: 22901812

90. Cone KR, Kronenberg ZN, Yandell M, Elde NC. Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J Virol. 2017; 91:e01428–16. https://doi.org/10.1128/JVI.01428-16 PMID: 27928012

91. Burke GR, Strand MR. Polydnaviruses of parasitic wasps: domestication of viruses to act as gene delivery vectors. Insects. 2012; 3:91–119. https://doi.org/10.3390/insects3010091 PMID: 26467950

92. Serbielle C, Dupas S, Perdereau E, Héricourt F, Dupuy C, Huguet E, et al. Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. BMC Evol Biol. 2012; 12:253. https://doi.org/10.1186/1471-2148-12-253 PMID: 23270369

93. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 2008; 42:165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119 PMID: 18983256

94. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012; 10:13–26. https://doi.org/10.1038/nrmicro2670 PMID: 22064560

95. Moran NA, Bennett GM. The tiniest tiny genomes. Annu Rev Microbiol. 2014; 68:195–215. https://doi.org/10.1146/annurev-micro-091213-112901 PMID: 24995872

96. Strand MR, Burke GR. Polydnaviruses: Evolution and Function. Curr Issues Mol Biol. 2020; 34:163–182. https://doi.org/10.21775/cimb.034.163 PMID: 31167960

97. Simmonds TJ, Carrillo D, Burke GR. Characterization of a venom gland-associated rhabdovirus in the parasitoid wasp *Diachasmimorpha longicaudata*. J Insect Physiol. 2016; 91–92:48–55. https://doi.org/10.1016/j.jinsphys.2016.06.009 PMID: 27374981

98. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013; 10:563–569. https://doi.org/10.1038/nmeth.2474 PMID: 23644548

99. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of *Drosophila*. Science. 2000; 287:2196–2204. https://doi.org/10.1126/science.287.5461.2196 PMID: 10731133

100. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

101. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

102. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153 PMID: 24642063

103. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012; 28:464–469. https://doi.org/10.1093/bioinformatics/btr703 PMID: 22199388

104. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011; 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195 PMID: 22039361

105. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:improvements in performance and usability. Mol Biol Evol. 2013; 30:772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

106. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29:644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440

107. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945

108. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

109. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562–578. https://doi.org/10.1038/nprot.2012.016 PMID: 22383036

110. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995; 57:289–300.

111. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 2004; 21:1095–1109. https://doi.org/10.1093/molbev/msh112 PMID: 15014145